Towards Modeling AI-based User Empowerment for Visual Big Data Analysis

Thoralf Reis[a], Sebastian Bruchhaus[a], Binh Vu[a], Marco X. Bornschlegl[a] and Matthias L. Hemmje[a]

[a] *University of Hagen, Faculty of Mathematics and Computer Science, 58097 Hagen, Germany*

Abstract

User empowerment for information systems has the objective to increase the system usability and the users' self-confidence. This can be achieved through offering additional information and adaptive user interfaces. Visual Big Data Analysis is an application domain that benefits from user empowerment since skilled personnel are rare and the infrastructure is expensive. The trending topic of AI is applied in industry and science to automate manual activities and to support human users. Based on the AI2VIS4BigData reference model, this work proposes an approach to utilize AI to empower users during their visual Big Data Analysis exploration journey. The proposal comprises a two step AI-based user empowerment concept for visual Big Data Analysis (insight extraction from Big Data and insight communication to end users) as well as a research roadmap.

Keywords

User Empowerment, AI, Big Data, Visualization, Big Data Analysis, AI2VIS4BigData, Use Cases, XAI

1. Introduction and Motivation

User empowerment for **Information Systems (IS)** has been discussed for some time now [1, 2, 3]. It has the objective to maximize value generation through a symbiotic relationship of the system user's intellectual potential and the system's capabilities. It comprises methods and principles that aim at increasing the users knowledge level and self-confidence to utilize as many of the opportunities an IS has to offer as possible [2]. Examples comprise empowering users to tailor IS **User Interfaces (UIs)** to their needs or to adapt the usage of the IS to be more goal-oriented and efficient. Knowledge, information, and psychological aspects [2] are the key of user empowerment: users of an IS are required to belief in their own skills and capabilities, they need to understand what the objective of utilizing the IS is, and how it can be influenced to what degree [2].

In general, the principles of user empowerment are not restricted to IS of any specific application domain. In [3] Bornschlegl et al. investigated end user empowerment for IS in the popular application domain of Big Data Analysis. This investigation make sense since there exist multiple challenges for this application domain [4] that are caused by the huge amount of data (high volume), the high data inflow (high velocity), and the high heterogeneity of the data (high variety) [5]; the challenges of high infrastructure costs through processing and storage of Big Data, a lack of accessibility due to the high dimensionality of the data, and insufficient testing and validation methods were mentioned as the most-important challenges for Big Data Analysis in an expert roundtable workshop in 2020 [4]. The high demand for skilled Big Data Analysis specialists [4] is another major challenge that makes designing an IS for that application domain even more difficult. User empowerment could address these challenges through adaptive UIs and intelligent communication of relevant information in order to empower Big Data Analysis user stereotypes, to use the system as efficiently and as successfully as possible.

IVIS4BigData is an existing standard for the design and implementation of IS for Big Data Analysis [6]. It is reference model with user empowerment as key component [6]. IVIS4BigData models the process of Big Data Analysis as four processing steps from data integration to the consumption of highly aggregated views and dashboards and empowers experts and end users to interact with the system and intermediate results over the whole process [6]. Therefore, the implementation of user empowerment is strongly connected to the principles of Fischer and Nakakoji's multifaceted architecture [3]. This architecture consists of three layers [1]: a domain knowledge layer, a design creation layer, and a feedback layer that connects domain knowledge layer and design layer [1]. This connection has the objective to empower users that specify and implement a design through critical reflections, case-based reasoning, and simulation based on the domain knowledge [1].

Fischer and Nakakoji explicitly decided not to design an expert system based on **Artifical Intelligence (AI)** that utilizes AI to replace human users but to utilize it to empower users for problem solving instead [1]. AI2VIS4BigData (Figure 1) is a reference model [7] that has the objective to continue this idea by describing how AI can be utilized for user-driven visual Big Data Analysis systems. This reference model is derived from IVIS4BigData and describes how AI models themselves pass through the IVIS4BigData processing steps (utilizing Big Data Analysis for designing AI models) as well as how deployed AI models could support the process of Big Data Analysis (deployed AI models for analytics, automation, and the UI) [7]. However, until now, AI2VIS4BigData lacks a detailed description how AI can be utilized for empowering Big Data Analysis users.

The objective of this short paper is now to introduce a research approach and a conceptual model that utilizes the ideas of Fischer and Nakakoji's multifaceted architecture [1] for the purpose of modeling AI-based user empowerment for visual Big Data Analysis IS. The next sections comprise a description of this conceptual model and a research roadmap before this short paper will conclude with an envisioned approach for its validation as well as a brief summary with an outlook on future research directions.
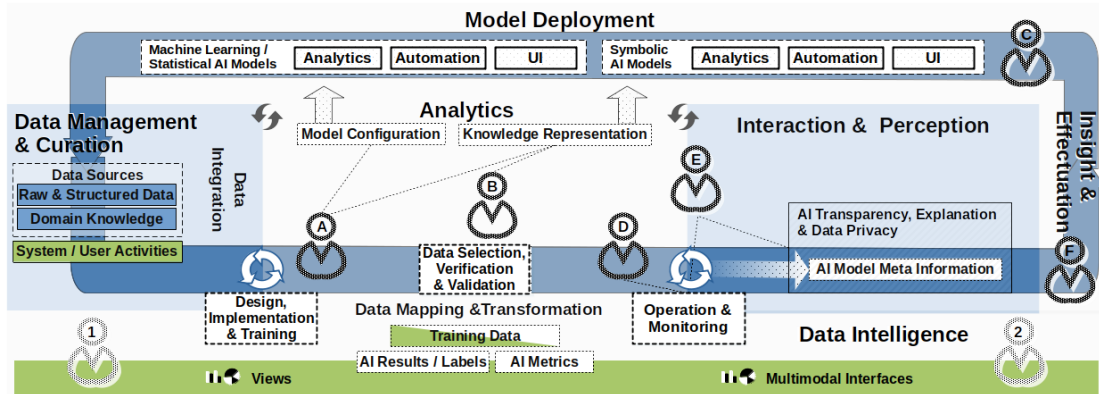
Figure 1: AI2VIS4BigData Reference Model for AI-based Visual Big Data Analysis [7]

2. Conceptual Modeling

There exist two main objectives of user empowerment for IS: the acquisition of additional domain knowledge and increasing the user's self-confidence in mastering the system's complexity [2]. This short paper proposes to interpret Fischer and Nakakoji's multifaceted architecture [1] for visual Big Data Analysis as follows in order to derive an approach for AI-based user empowerment: the goal of design creation shall be the formulation and execution of a fruitful Big Data Analysis that fulfills the user's information demand; this design creation requires the user to be self-confident and informed enough about potential decisions (specification) and to actually realize the Big Data Analysis (construction); insight about the data (semantics base) as well as experiences about successful or inefficient workflows (catalog and argumentation base) support the user to be as self-confident as possible. The resulting interpretation is visualized in Figure 2 and divided into the two steps of Big Data insight extraction (step 1) and insight communication (step 2). This interpretation is in agreement with Norman and Draper's user-centered system design approach in which they *"focus on the substantive details and origins of the mental models users construct for how computational systems work"* [8] in order to bridge the gap between user and the system. They proposed three criteria to be relevant for users to understand IS: *"internal coherence, validity, and integration of available and new knowledge"* [8].

Following the interpretation in Figure 2 and the criteria of integration new knowledge for Norman and Draper's user-centered system design [8], a potential way of introducing AI-based user empowerment for the different user stereotypes in visual Big Data Analysis is to utilize AI to extract the required Big Data insights and to transport them by utilizing rules (predefined or symbolic AI) to the user. These insight can either be transported via direct visual guidance (e.g., color highlighting of relevant data artifacts) or via indirectly emphasizing the interaction with the system (e.g., reordering a context menu according to data characteristics). The user story in Figure 3 exemplarily visualizes this process.
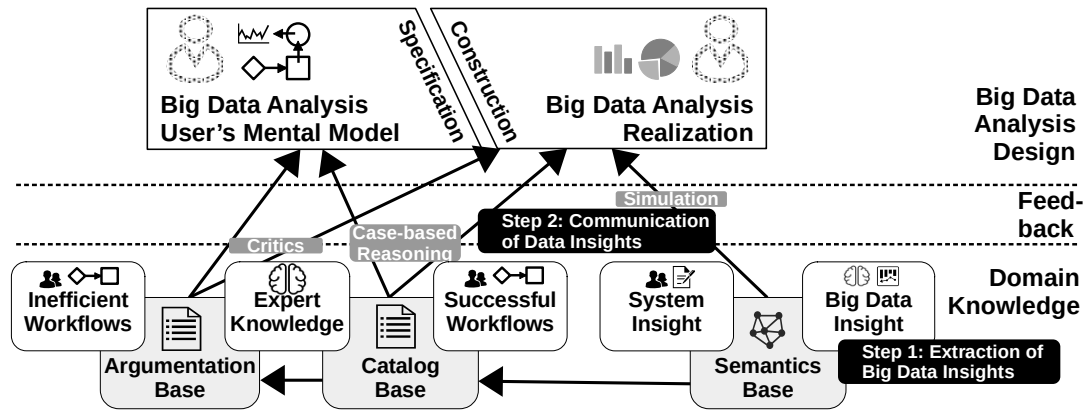
Figure 2: Multifaceted Architecture [1] Interpretation for User Empowerment in Visual Big Data Analysis
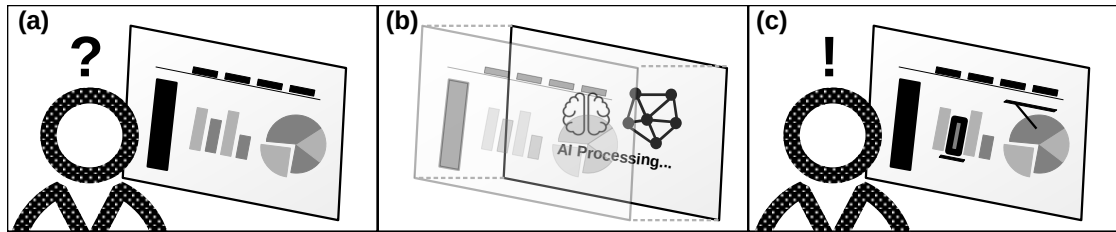

Figure 3: User Story for AI-based User Empowerment in Visual Big Data Analysis

Figure 3 captures a visual Big Data Analysis IS and a user that utilizes it at three points in time. Starting with (a), an uncertain user visually explores Big Data by analyzing graphical representations of the data in a traditional way. Without requiring further user input, the IS executes predefined AI models in (b) that extract features and insight from the data. In (c), the user is informed about this additional knowledge about the data through visual highlighting of the data visualization and gains self-confidence for further IS interaction. Thus user empowerment in context of this work means to enhance the user's self-confidence to interact with an IS in a sophisticated manner that lets him/her exploit the system's full potential.

In order to model the two steps of extraction of information from Big Data (step 1) and of transporting this information to the user (step 2) in more detail, specific and practical use cases need to be derived and integrated within a UI. Figure 4 translates the user story from Figure 3 into a wireframe based on the IVIS4BigData UI implementation [6]. The wireframe visualizes the exemplary use case of AI-based data hotspot detection. It shows the IVIS4BigData UI in which multiple integrated and analyzed data collections are presented to the user. Within this scene, the user can select one or multiple data collections
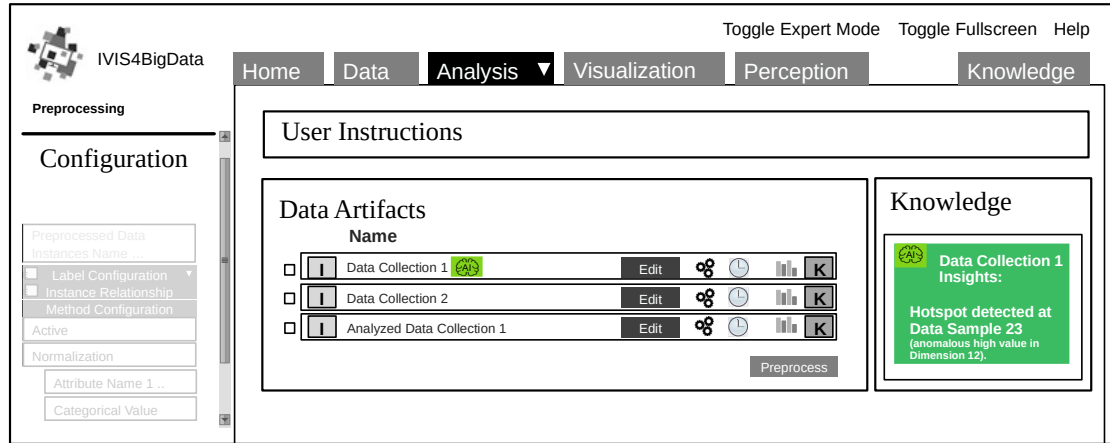
Figure 4: Wireframe of IVIS4BigData [6] and AI2VIS4BigData UI establishing AI-based User Empowerment

and apply an analysis method or visualize the data. In this example, AI2VIS4BigData executes an AI model to detect hotspots within these data collections. After application of this model and extraction of hotspot information, the identified hotspot is communicated to the user via a pre-determined area for knowledge on the data artifacts located at the right side of the wireframe.

3. Research Roadmap

The application, implementation, and validation of the derived conceptual model requires specific application scenarios and target user groups, concrete ideas how the information can be transported to these users, and a comprehensive evaluation strategy. For this purpose, this paper proposes a three step research roadmap that is explained within the following subsections.

3.1. Identifying User Empowering Use Cases

Beyond detecting hotspots within the data (Figure 4), there exists various application scenarios for empowering Big Data Analysis users. The identification of such scenarios can either be carried out in a theoretical and deductive or in a practical and inductive way. This short paper proposes to combine both strategies to identify these scenarios and to design further UI wireframes:

- Deductive approach: Identifying manual activities in visual Big Data Analysis through examination of all manual activities that are required to go through all processing steps of the theoretical IVIS4BigData reference model
- Inductive approach: Assessment of practical and relevant user challenges through conducting a literature research or an expert survey

Since specification and implementations of these use cases will require further information, the proposed approach contains two further steps that enable putting the use cases into context:

- Derivation of a use case framework consisting of all relevant elements from AI2VIS-4BigData and Fischer Nakakoji's multifaceted architecture through extending the existing IVIS4BigData use case framework [3]
- Contextualizing the use cases through modeling the relationships among each other as well as the relationships to reference model elements within this use case framework

3.2. User Empowerment with Explainable AI

The potential of user empowerment for visual Big Data Analysis as motivated in the previous sections goes beyond enhancing the ability to draw conclusions from data sections. AI systems should also qualify a certain degree of confidence in their results. The collaborative approach to data processing with artificial and human actors requires trust and "digital empathy" [9, 10]. The intuitive concept of trustworthiness relies on predictability and hence some form of **Explainable AI (XAI)** [11].

A prevalent hypothesis concerning XAI suggests a trade-off between interpretability and accuracy of models. **Deep Learning (DL)** is an example of "black box" models under this assumption. Yet model-agnostic explanation methods like, e.g., LIME and SHAP and those specific to DL like, e.g., GradCAM have put this notion somewhat into perspective [12, 13, 14]. Different explanatory techniques can be combined to enhance their overall efficacy [15]. Even more principled approaches to XAI like, e.g., Bayesian learning have been incorporated into **Machine Learning (ML)** frameworks such as TensorFlow [16].

Enhanced transparency usually causes computational cost and cognitive load for the user, but not all automated decisions require justifications. The users of ML models therefore ought to be able to choose the appropriate kind of explanations as well as the mode of their mediation. Decision support in this regard is itself a candidate for automation by recommender systems. Data quality monitoring ought to be considered in this context as well, because the performance of ML models corresponds to the quality of its training data.

There are mainly two ways for conveying explanations, i.e. textual [17] and visual. The latter overlaps extensively with statistical visualizations, e.g., the familiar line of best fit of a simple linear regression model or graph plots of Bayes belief networks.

This work proposes that user stereotypes should be offered appropriate ML models and explanation techniques along each step of AI2VIS4BigData. Context sensitive dialogue systems are an obvious method for this. They should not interrupt the workflow but rather offer reasonable defaults. Explanations of automated decisions can then be generated in parallel to the actual ML process. Recommendation systems for the desired form of XAI can help to minimize the cognitive load on user stereotypes. Visual explanations are to be overlayed above the actual visualizations in an optional graphical layer. These textual and visual explanations shall then complement the text in Figure 4 for the purpose of

user empowerment. A linear regression plot is a simplistic example in this context that lets the user check the trained models' performance at a glance.

3.3. Validation of the Conceptual Model

Since user empowerment for IS is strongly related to psychological aspects [2] and the proposed conceptual model depends on the user's mental model [8], a validation strategy is challenging. In a nutshell, there are to potential ways to validate the proposed approach for AI-based user empowerment:

1. Passive, observatory validation of the user's mental model according to Norman and Draper [8] through, e.g., measurement of the required time with and without additional knowledge until the user successfully imagined how the IS can be utilized for a certain Big Data Analysis task.
2. Active, interrogating validation of a certain user while being presented multiple wireframes of the system.

This paper proposes to follow the second option in form of a cognitive walkthrough, an evaluation technique that focuses on *"how well [...] first-time use without formal training"* [18] of a system is. A cognitive walkthrough can be conducted with wireframes and does not require a full implementation of the system. During this validation, an UI or domain expert user is asked to fulfill a set of tasks for which he/she has to interact with the system. The validation reaches a quantitative result through comparing the user's action with *"a list of the correct actions required to complete each of these tasks"* [18] as well as qualitative results through interviewing the test user afterwards regarding perceived challenges. For this, all IVIS4BigData processing steps can be passed through with multiple wireframes for AI-based user empowerment examples.

4. Summary and Outlook

This short paper presents a research approach that enables modeling AI-based user empowerment for visual Big Data Analysis through AI-based Big Data insight extraction and communication of this insight to expert and end user stereotypes. This approach was derived through interpretation of Fischer and Nakakoji's multifaceted architecture [1] and is based on the AI2VIS4BigData reference model. In addition, this paper introduced a research roadmap of immediate next research objectives in order to employ the conceptual model.

This research roadmap comprises the derivation of AI-based use cases for Big Data insight extraction, the definition of an AI2VIS4BigData use case framework, the instantiation of this use case framework to model the use cases, the investigation of XAI's potential for empowering end users, and the validation of the conceptual model for practical use cases. In addition the outlook comprises detailed specifications and prototypical implementations of the conceptual model as well as looking into the differentiation between inefficient and successful workflows. The latter enables a comprehensive application of the multifaceted architecture within an IS.

References

[1] G. Fischer, K. Nakakoji, Beyond the macho approach of artificial intelligence: empower human designers—do not replace them, Knowledge-Based Systems 5 (1992) 15–30.

[2] H.-W. Kim, S. Gupta, A user empowerment approach to information systems infusion, IEEE Transactions on Engineering Management 61 (2014) 656–668.

[3] M. X. Bornschlegl, K. Berwind, M. L. Hemmje, Modeling end user empowerment in big data applications, 26th International Conference on Software Engineering and Data Engineering, SEDE 2017 (2017) 47–53.

[4] T. Reis, M. X. Bornschlegl, M. L. Hemmje, AI2VIS4BigData: A Reference Model for AI-based Big Data Analysis and Visualization, in: Advanced Visual Interfaces. Supporting Artificial Intelligence and Big Data Applications, volume 12585 of *Lecture Notes in Computer Science*, Springer International Publishing, 2021, pp. 1–18. doi:`10.1007/978-3-030-68007-7\_7`.

[5] D. Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical Report, META Group, 2001.

[6] M. X. Bornschlegl, Advanced Visual Interfaces Supporting Distributed Cloud-Based Big Data Analysis, Dissertation, University of Hagen, 2019.

[7] T. Reis, M. X. Bornschlegl, M. L. Hemmje, Towards a Reference Model for Artificial Intelligence Supporting Big Data Analysis, To appear in: Advances in Data Science and Information Engineering -Proceedings of the 2020 International Conference on Data Science (ICDATA'20) (2021).

[8] R. Pea, User Centred System Design-New Perspectives on Human/Computer Interaction, Journal of Educational Computing Research 3 (1987).

[9] R. Bond, F. Engel, M. Fuchs, M. Hemmje, P. M. Kevitt, M. McTear, M. Mulvenna, P. Walsh, H. J. Zheng, Digital empathy secures Frankenstein's monster, CEUR Workshop Proceedings 2348 (2019) 335–349.

[10] P. A. Hancock, D. R. Billings, K. E. Schaefer, Can you trust your robot?, Ergonomics in Design: The Quarterly of Human Factors Applications 19 (2011) 24—-29. doi:`10.1177/1064804611415045`.

[11] D. Gunning, D. Aha, Darpa's explainable artificial intelligence (xai) program, AI Magazine 40 (2019) 44–58. doi:`10.1609/aimag.v40i2.2850`.

[12] M. T. Ribeiro, S. Singh, C. Guestrin, "why should i trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144. doi:`10.1145/2939672.2939778`.

[13] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774.

[14] A. Grabska-Barwińska, Measuring and improving the quality of visual explanations, CoRR (2020). `arXiv:2003.08774`.

[15] P. Biecek, T. Burzykowski, Explanatory Model Analysis, CRC PRESS, 2021.

[16] D. Tran, M. W. Hoffman, D. Moore, C. Suter, S. Vasudevan, A. Radul, Simple, distributed, and accelerated probabilistic programming, in: Advances in Neural Information Processing Systems, volume 31, Curran Associates, Inc., 2018, pp. 7598–7609.

[17] T. Eljasik-Swoboda, F. Engel, M. Hemmje, Explainable and transferrable text categorization, Data Management Technologies and Applications (2020) 1–22. doi:`10.1007/978-3-030-54595-6_1`.

[18] J. Rieman, M. Franzke, D. Redmiles, Usability evaluation with the cognitive walkthrough, in: Conference companion on Human factors in computing systems, ACM, 1995, pp. 387–388.